



Data Recovery Effort of Nimbus era Observations by the NASA GES DISC

James Johnson^{1,2}, Ed Esfandiari^{1,2}, Emily Zamkoff^{1,3},
Irina Gerasimov^{1,2}, Atheer Al-Jazrawi^{1,3} and Gary Alcott¹

1. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA GSFC

2. ADNET Systems, Inc.

3. Telophase

5th International Conference on Reanalysis 2017
14 November 2017

<https://disc.gsfc.nasa.gov>



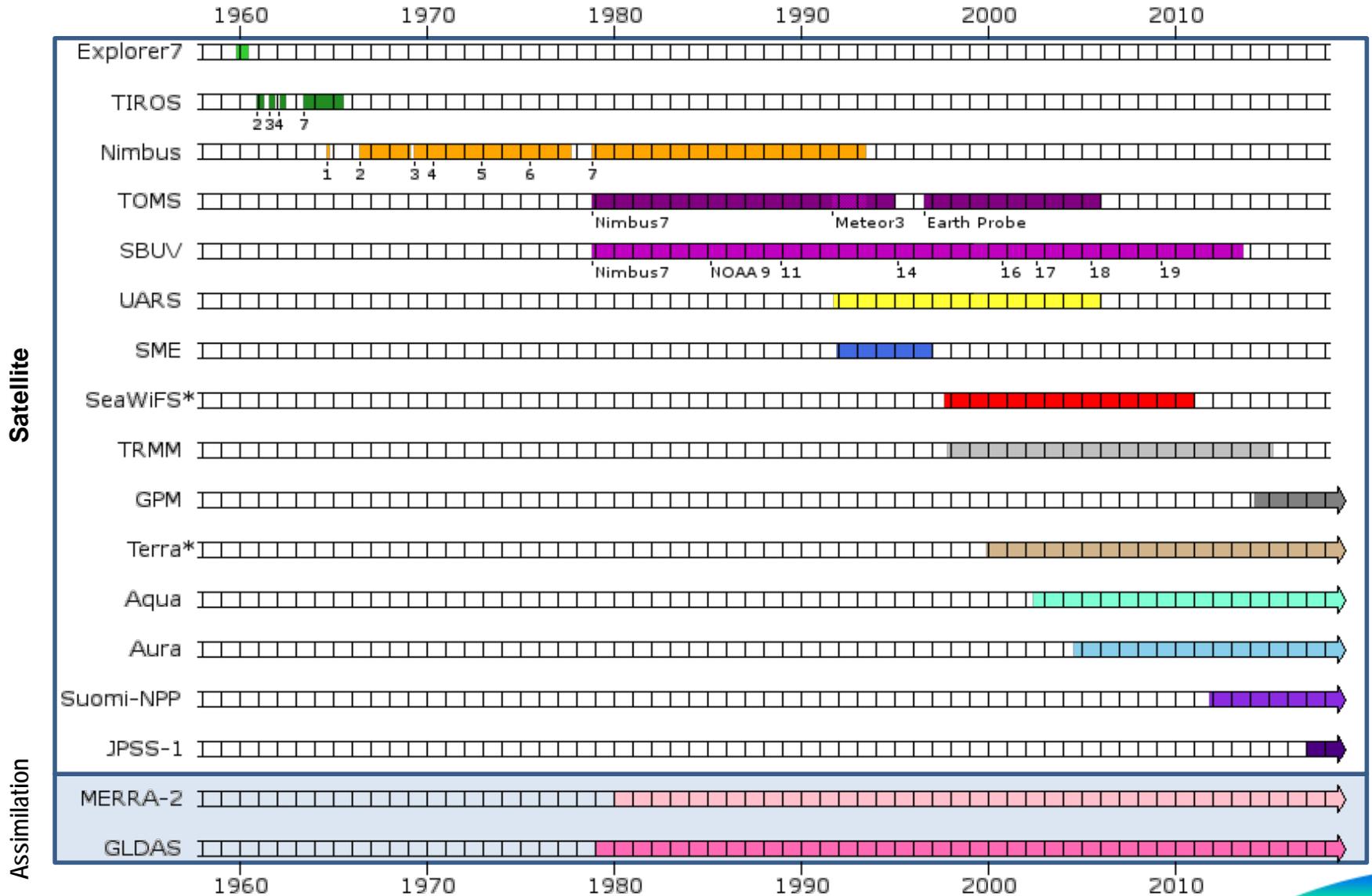
Introduction

At end of mission data went to NASA's National Space Science Data Center (*NSSDC*), and from there to the National Archives Federal Record Center (*FRC*)

- **Earth Science Data Recovery Task:**
 - Preserve Nimbus era data written on 7- and 9-track tapes, 3480 cartridges, film imagery, and supporting documentation
 - Make data accessible online to the scientific community
 - Free up space occupied by bulky media and need for climate controlled warehouse
 - Funded by NASA's Earth Science Data and Information System (*ESDIS*) project
 - Implemented and coordinated by NASA's *GES DISC*
- **Data Recovery Issues:**
 - Fragile media dating back to the early 1960s
 - Lack of useful and applicable documentation
 - Knowledgeable personnel for consultation no longer available
 - Data quality is lacking
 - Time consuming, often requiring manual intervention
 - Non-existent metadata



~60 Years of Earth Data at GES DISC

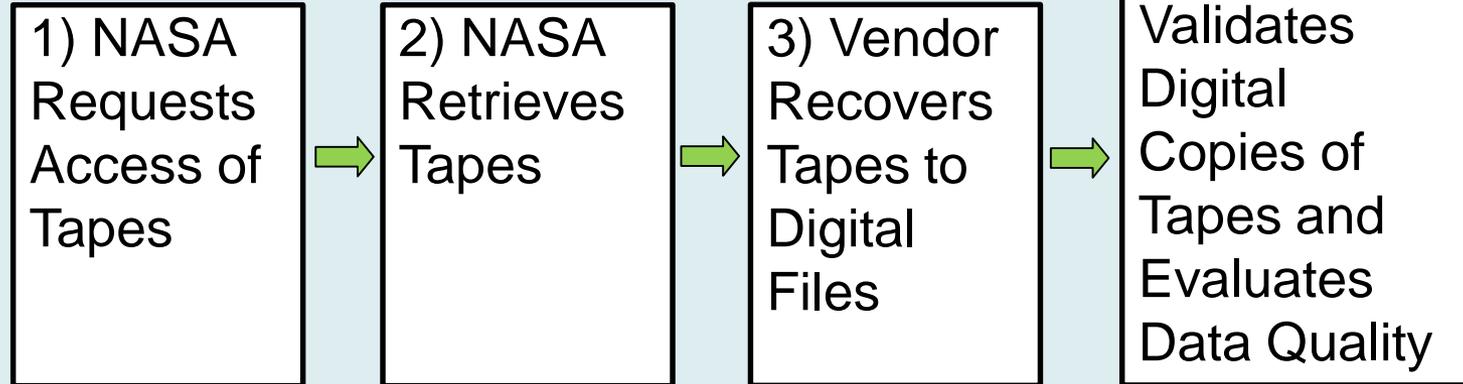


* Explorable through GES DISC Giovanni Visualization Service



Recovery Process

Data Recovery



Data Processed





Extract Data Files from Tape File

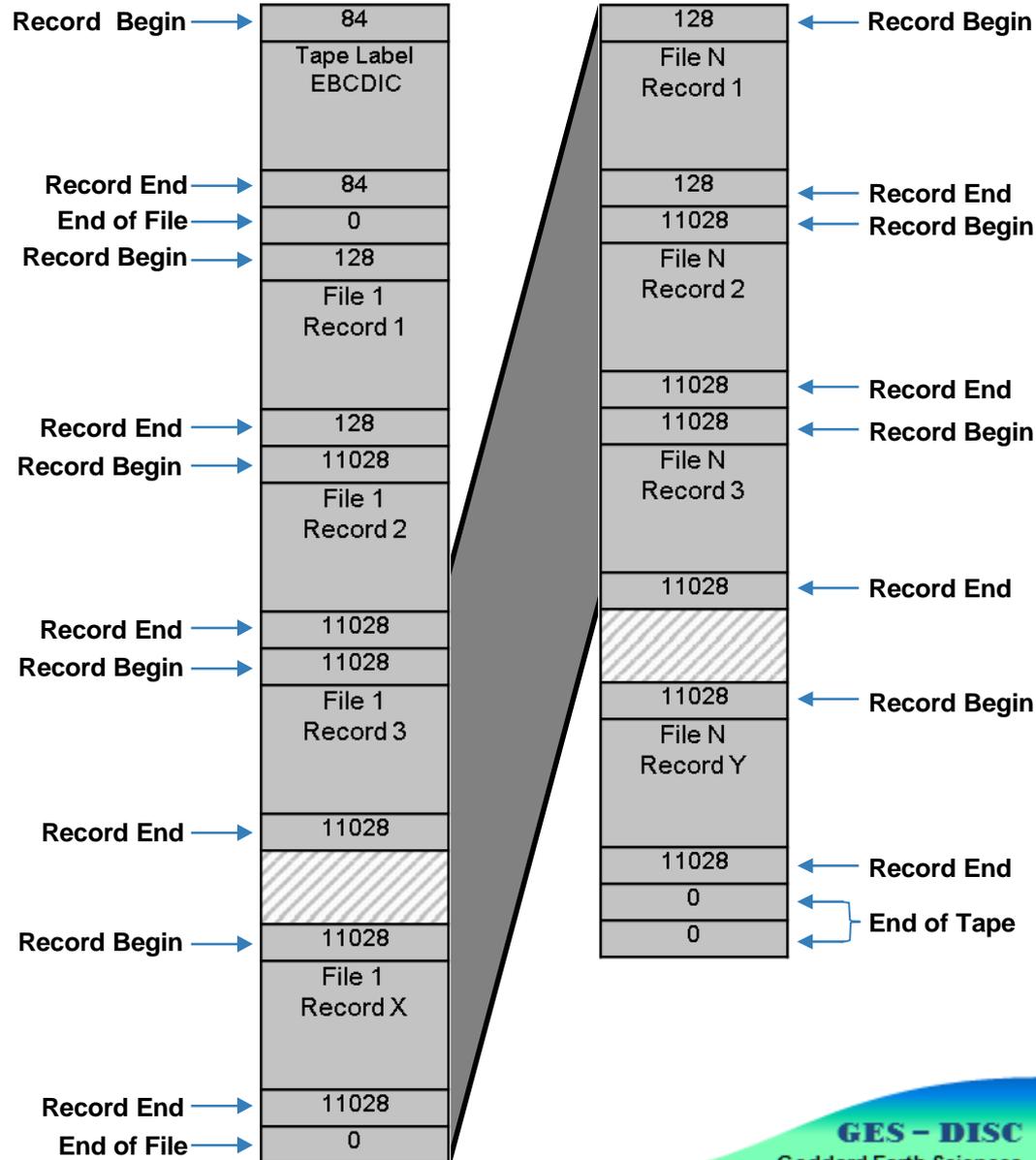
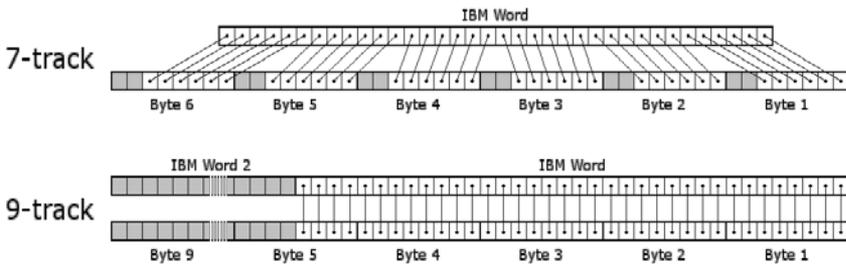
- In the Nimbus era, each experiment team designed their own unique file format, limits software reuse
- No concept of granule level metadata, this has to be extracted from each granule or data file and created new
- Data originally written on outdated IBM-360 machines:
 - use 36-bit or 32-bit words
 - use IBM integer, floats and characters (EBC not ASCII)
- Files have no names, GES DISC creates names based on metadata: experiment, date, orbit and tape number
- Backup tapes must be reviewed individually and compared with primary tape for any missing data files



Nimbus 2 HRIR TAP File Format

TAP Format Header: 32-bit integer
 bit 0-30: length of record in bytes
 bit 31: 0 = good record, 1 = bad

Reconstructing the original data:





HRIR, MRIR and THIR Data Files

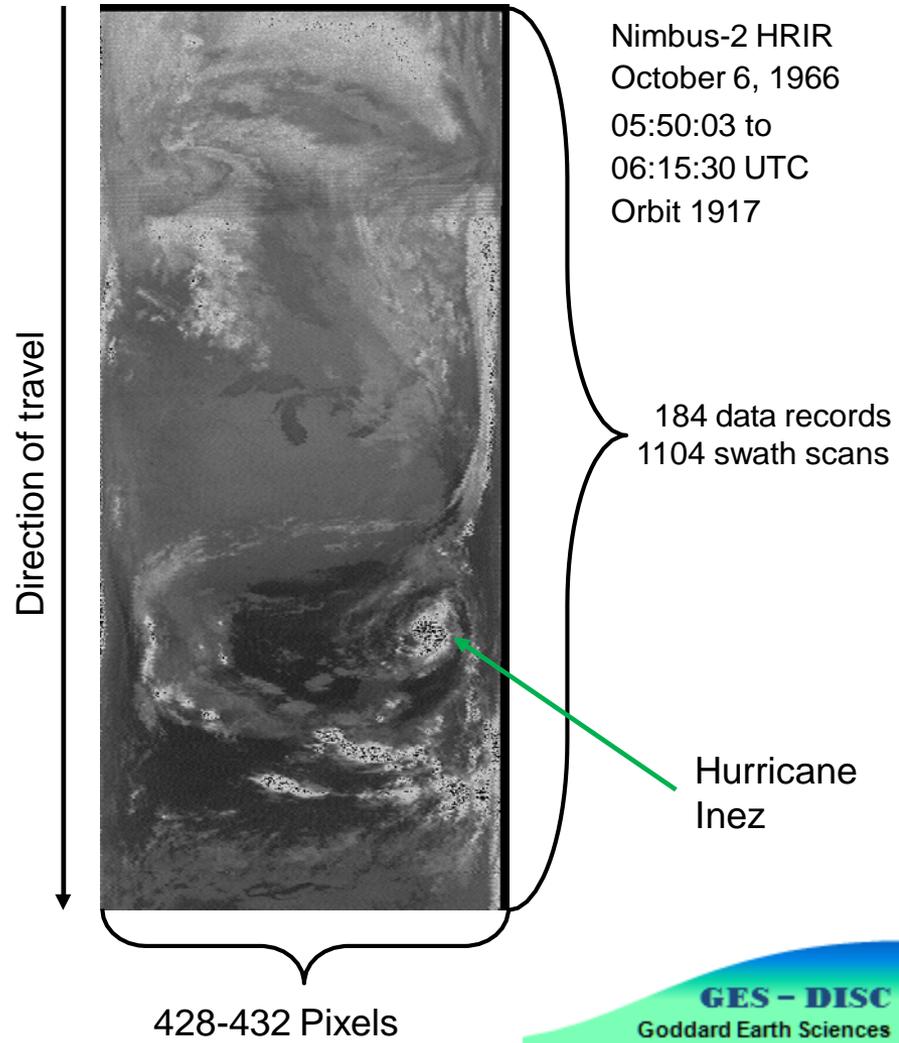
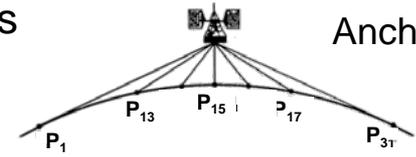
- Data originally created on IBM-360 using 36-bit words
- Data packed in either 6 x 6-bit or 4½ x 8-bit bytes
- The original file structure is preserved

128
Documentation Record
128
11028
Data Record 1
11028
11028
Data Record 2
11028
11028
Data Record 3
11028
11028
Data Record X
11028
0

Documentation Record
 Start Date and Time
 End Date and Time
 Orbit Retrieval Number
 Number of location anchor points
 (31 though typically 29 used)
 Swath size (in words)
 Number of swaths (6 per record)

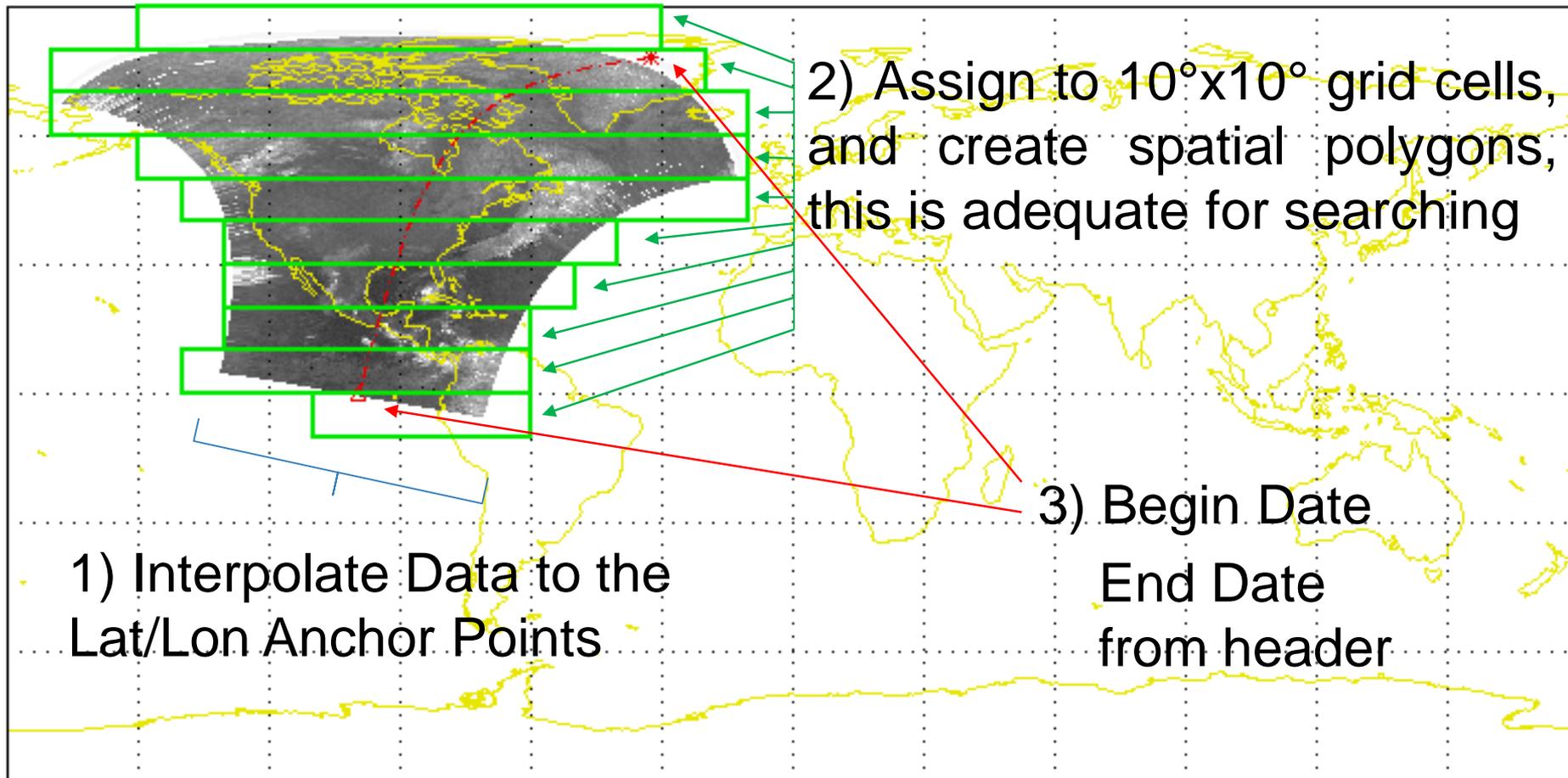
Data Record Header
 Date and time
 Pitch, yaw, roll errors
 Hardware status
 Nadir angles for anchor points (31)

Swath (repeats 6 times)
 Start time (seconds)
 Channel (for MRIR)
 Number of data points in swath
 Sub-satellite lat and lon
 Anchor point lat and lon (31)
 Instrument status flag
 Brightness Temperatures (~430)





The File-Level Metadata



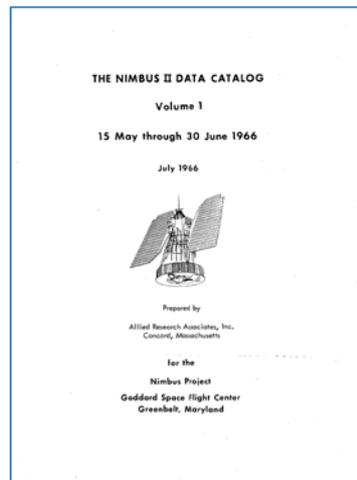
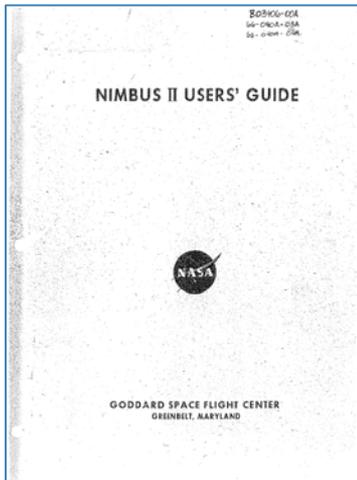
4) Extract Orbit from header
(actually retrieval orbit)

5) Add Recovery Contractor
QA metadata



Documentation

- GES DISC web site contains directory of Nimbus data products, and supporting documentation: User's Guides, Data Catalogs, and READMEs.
- Inventory of all tapes and files also ingested.
- Some Hardcopies must be scanned.





Data Recovery Issues

- Bookkeeping
- Documentation
- Media
- Data Processing



Bookkeeping Issues

- Data from Unrelated Mission
 - Operator Error not Rewinding the Tape?
 - Operator Attempt to Maximize use of Limited Resource?
- Incorrect Tape Label
 - Missing Label
 - Hard to Read Handwriting
 - Reused Tape but not Relabeld
 - Incorrect Information (e.g. collection, date, orbit, format)



Documentation Issues

- **Lack of Useful Documents**
 - Hard to locate documents to correctly describe the data being recovered
- **Documents Sometimes Do not Reflect Data Structure**
 - Earlier version of document does not reflect final data format
 - Different modes/anomalies understood at the time of the mission but not reflected in final archived document



Media Issues

- **Sticky Tape**
 - Common problem (sticky-shed syndrome) due to excessive moisture during storage; tape must be carefully baked before reading
- **Fragile Media**
 - Stress may cause tape to stretch, tear, or scratch making data unreadable
 - Coating worn from substrate making data unrecoverable
- **Broken Reel**
 - If broken, contractor may be able to reassemble the hub
 - If not, tape may be unrecoverable if tape cannot be transferred to new reel
- **Missing Begin or End of Tape Marker**
 - In a few cases, contractor was able to locate and attach new marker



Data Processing Issues

- Detect if Data is from 7-track or 9-track Tape
 - Convert 7-bits from 7-track tape (*6 bit plus parity*) to 8 bits:
 - Add extra bit
 - To extract the original 36-bit IBM word, read **6 8-bit bytes**, ignore 6th and 7th bits of each byte and combine the remaining bits
 - Convert 9-bits from 9-track tape (*8 bit plus parity*) to 8 bits:
 - Drop parity bit
 - To extract the original 36-bit IBM word, read **4½ 8-bit bytes** and then combine the bits
- Determine Endianness
 - Usually *big-endian*, sometimes *little-endian*, modify code accordingly
- Multiple Tape Formats in a Collection
 - 7-track, 9-track, or even 3480 cartridges
- Missing or Multiple Tape Label Records
 - Common problem, code modified to detect/skip these



Data Processing Issues (cont.)

- Missing or Extra Orbit Records
 - Orbit info often used in filenames, typically handled manually
- Missing End-of-File and/or End-of-Tape
 - Due to tape degradation or error when tape was originally written
- Invalid Record Lengths (*frequent for older data*)
- Files from Different Collection on the Same Tape
- Tapes not Rewound when Originally Written
 - Many unrelated bytes of data before first Nimbus data found
- Corrupt Tapes (*nothing recoverable < 1%*)
- Unknown File Format
 - Lack of or due to poor documentation (*requires guess work and time consuming*)
- Duplicate Data Files
 - Ensure code doesn't overwrite



Nimbus Dataset Status

Nimbus			1	2	3	4	5	6	7
Infrared Imagers	HRIR	High Resolution Infrared Radiometer	Public	Public	Public				
	MRIR	Medium Resolution Infrared Radiometer		Public	Public				
	THIR	Temperature and Humidity Infrared Radiometer				Public	Public	Public	Public
Microwave Imagers	ESMR	Electronic Scanning Microwave Radiometer					Public		
	SMMR	Scanning Multispectral Microwave Radiometer							TBD
Infrared Sounders	IRIS	Infrared Interferometer Spectrometer			Missing	Public			
	SIRS	Satellite Infrared Spectrometer			Processed	Processed			
	SCR	Selective Chopper Radiometer				x	x		
	ITPR	Infrared Temperature Profile Radiometer					Recovered		
	HIRS	High Resolution Infrared Sounder						Public	
	LRIR	Limb Radiance Inversion Radiometer						Processed	
	PMR	Pressure Modulated Radiometer						Public	
	LIMS	Limb Infrared Monitor of the Stratosphere							x
	SAMS	Stratospheric and Mesospheric Sounder							x
Microwave Sounders	NEMS	Nimbus-E Microwave Sounder					Recovered		
	SCAMS	Scanning Microwave Spectrometer						Public	
Ultraviolet Sensors	BUV	Backscatter Ultraviolet Spectrometer				Public			
	SBUV	Solar Backscatter Ultraviolet Spectrometer							Public
	TOMS	Total Ozone Mapping Spectrometer							Public
Other	SCMR	Surface Composition Mapping Radiometer					Recovered		

Public
 Processed
 Recovered
 Missing
 TBD
 x = Add'l tape data to be recovered

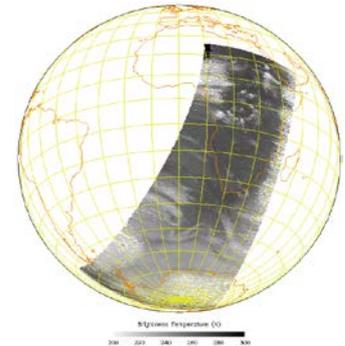
NOTE: AVCS + ITPS + SMMR Snow/Ice to NSSDC; ERB + SAM-II to ASDC



Conclusion

- This is tedious work!
- Important to preserve the data, otherwise lost forever!!!
- No common format makes each product unique
 - *limits software reuse*
- File formats sometimes deviate from documentation
- Corrupted records and data make extraction hard
- Corrupted tapes makes data unrecoverable
- See <https://disc.gsfc.nasa.gov> for access to the data, documentation, and for more information
- **Reference:** Khayat, M., Kempler, S., “Life Cycle Management Considerations of Remotely Sensed Geospatial Data and Documentation for Long Term Preservation,” 2017, <https://ntrs.nasa.gov/search.jsp?R=20160002963>

First data from
Nimbus-1 HRIR
1964/08/29
Orbit 23



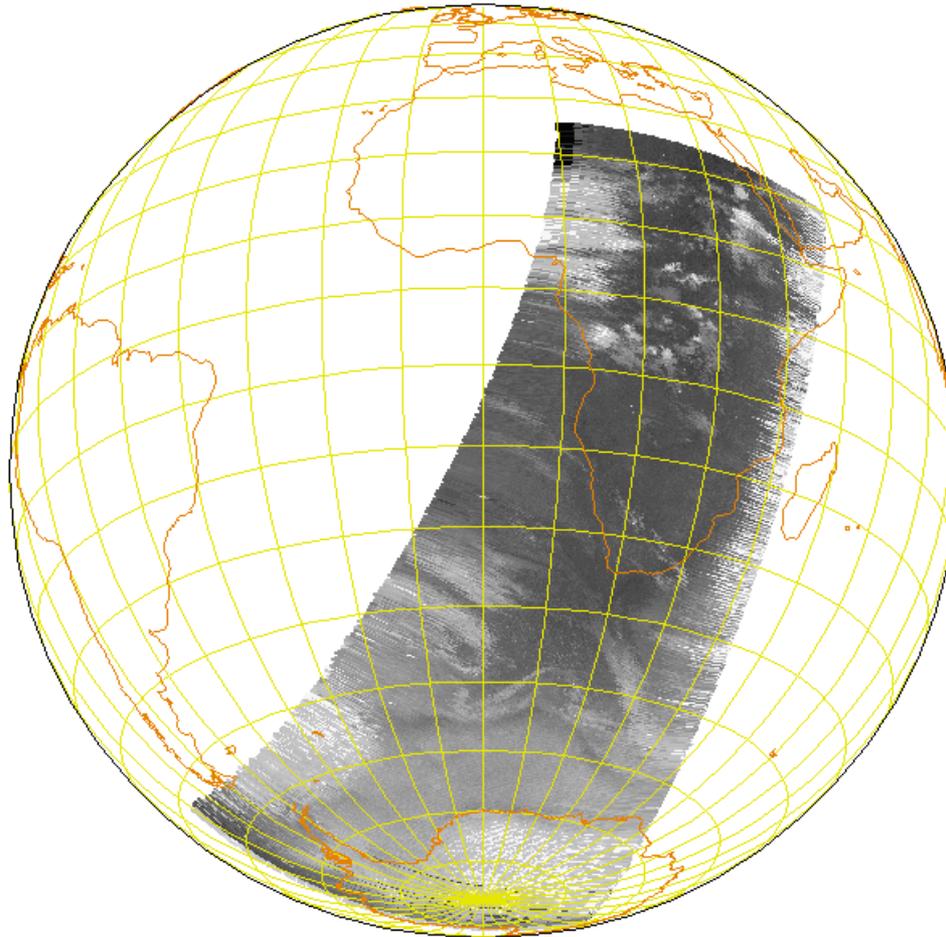


Extra



First Nimbus-1 HRIR Data File

Nimbus-1 HRIR - Date: 1964/08/29 21:46:15Z - Orbit: 23



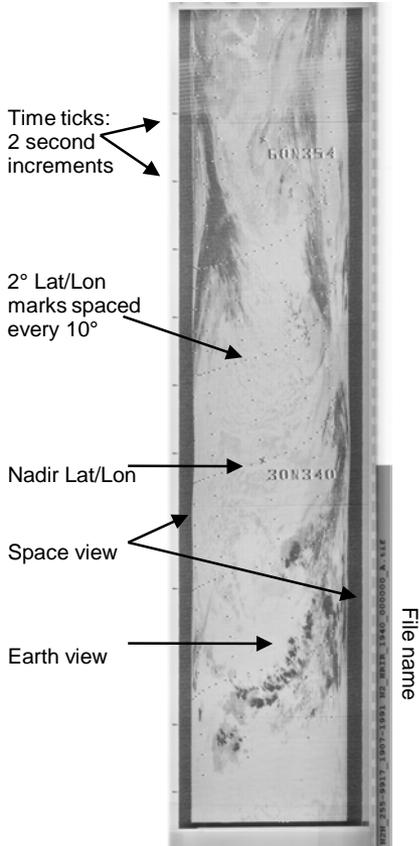
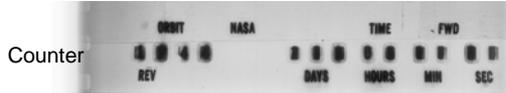
Brightness Temperature (K)



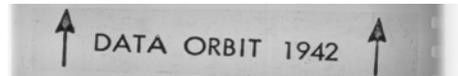
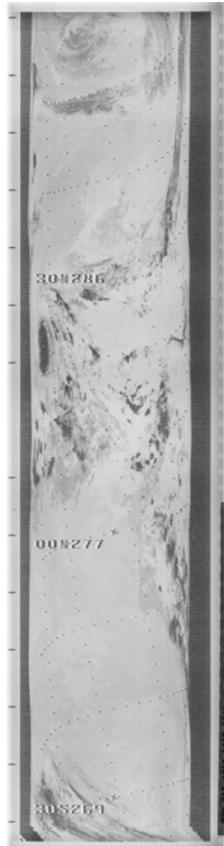


HRIR Film Strip Data Problems

Unreadable



Digits Wrong



Orbit off by 1



- About 1 week of images were tarred together and archived



Instruments Flown on Nimbus

Nimbus 1 – 1964/08/28	Nimbus 4 – 1970-04-08	Nimbus 6 – 1975-06-12
Advanced Vidicon Camera System (AVCS)	Image Dissector Camera System (IDCS)	Temperature and Humidity Infrared Radiometer (THIR)
Automatic Picture Transmission (APT) System	Temperature and Humidity Infrared Radiometer (THIR)	Electrically Scanning Microwave Radiometer (ESMR)
High-Resolution Infrared Radiometer (HRIR)	Infrared Interferometer Spectrometer (IRIS)	Limb Radiance Inversion Radiometer (LRIR)
	Satellite Infrared Spectrometer (SIRS)	Earth Radiation Budget (ERB)
Nimbus 2 – 1966/05/15	Backscatter Ultraviolet (BUV) Spectrometer	Pressure Modulated Radiometer (PMR)
Advanced Vidicon Camera System (AVCS)	Filter Wedge Spectrometer (FWS)	High Resolution Infrared Radiation Sounder (HIRS)
Automatic Picture Transmission (APT) System	Selective Chopper Radiometer (SCR)	Scanning Microwave Spectrometer (SCAMS)
High-Resolution Infrared Radiometer (HRIR)	Monitor of Ultraviolet Solar Energy (MUSE)	
Medium-Resolution Infrared Radiometer (MRIR)		Nimbus 7 – 1978-10-24
		Temperature and Humidity Infrared Radiometer (THIR)
Nimbus 3 – 1969-04-14	Nimbus 5 – 1972-12-11	Coastal Zone Color Scanner (CZCS)
Image Dissector Camera System (IDCS)	Temperature and Humidity Infrared Radiometer (THIR)	Stratospheric and Mesospheric Sounder (SAMS)
High-Resolution Infrared Radiometer (HRIR)	Infrared Temperature Profile Radiometer (ITPR)	Limb Infrared Monitor of the Stratosphere (LIMS)
Medium-Resolution Infrared Radiometer (MRIR)	Electrically Scanning Microwave Radiometer (ESMR)	Earth Radiation Budget (ERB)
Infrared Interferometer Spectrometer (IRIS)	Nimbus-E Microwave Spectrometer (NEMS)	Stratospheric Aerosol Measurement II (SAM-II)
Satellite Infrared Spectrometer (SIRS)	Selective Chopper Radiometer (SCR)	Scanning Multispectral Microwave Radiometer (SMMR)
Monitor of Ultraviolet Solar Energy (MUSE)	Surface Composition Mapping Radiometer (SCMR)	Solar Backscatter Ultraviolet Spectrometer (SBUV)
		Total Ozone Mapping Spectrometer (TOMS)



The Collection-Level Metadata

- GES DISC uses the GCMD DIF10 for storing collection level metadata in the Common Metadata Repository (CMR), can be used by other discovery tools.
- Allows for common looking landing pages across the GES DISC site.
- Metadata is populated from the Nimbus User's Guides and information available from NSSDC web pages.



Pre-Nimbus Datasets

Explorer 7

- 1959-11-15 to 1960-05-24 Thermal Radiation Experiment

TIROS 2

- 1960-11-23 to 1961-04-13 Scanning Radiometer

TIROS 3

- 1961-07-12 to 1961-10-20 Scanning Radiometer
- 1961-07-12 to 1961-10-20 Low-Resolution Omnidirectional Radiometer

TIROS 4

- 1962-02-08 to 1962-06-30 Scanning Radiometer
- 1962-02-08 to 1962-06-28 Low-Resolution Omnidirectional Radiometer

TIROS 7

- 1963-06-19 to 1965-06-19 Scanning Radiometer
- 1963-06-19 to 1963-08-29 Low-Resolution Omnidirectional Radiometer



Other Datasets

Solar Mesosphere Explorer (SME)

- 1981-12-16 to 1986-12-18 UV Ozone

EOLE 1 / Cooperative Application Satellite (CAS) 1

- 1971-08-27 to 1972-07-04 Upper Atmosphere Winds and Weather Data Relay System

Synchronous Meteorological Satellite (SMS) 2

- 1975-02-17 to 1975-08-28 VISSR

Geostationary Earth Observing Satellite (GEOS) 1

- 1975-02-17 to 1987-08-27 VISSR

Defense Meteorological Satellite Program (DMSP)

- 1977-03-25 to 1980-02-16 Multi-Channel Filter Radiometer (SSH)

Applications Technology Satellite (ATS) 6

- 1974-06-17 to 1974-08-30 Geosynchronous Very High Resolution Radiometer

Geodetic Earth Observing Satellite (GEOS)

- 1968-03-18 to 1968-07-25 GEOS-2 Optical Beacon Data
- 1975-04-09 to 1975-12-23 GEOS-3 Satellite-to-Satellite Tracking Data

Various Space Shuttle Data

- STS-2, STS-41G, STS-51B



Earth Science Data Recovery Accomplishments

Accomplishments of Earth Science Data Recovery for FY17

- Team members were co-authors on a paper, “Recent Advances in Satellite Data Rescue,” that was published in the Bulletin of the American Meteorological Society’s July 2017 issue: James Johnson, Asghar Esfandiari, Irina Gerasimov, Emily Zamkoff, Atheer Al-Jazrawi, <http://journals.ametsoc.org/doi/full/10.1175/BAMS-D-15-00194.1>.
- Validated the Nimbus 7 THIR Cloud Data for SBUV/TOMS (BCLT) data. 1516 files were ingested into archive and made available to the public.
- Sending film data (Nimbus, ATS-6, STS-41G, GOES, SMS) to USGS EROS for recovery.
- Received the final delivery from JBI and closed out all the purchase orders. JBI returned the 7-track tapes that they were unable process.
- Worked with the University of Wisconsin for the recovery of the SMS 9-track tapes.
- Work with the Atmospheric Trace Molecule Spectroscopy (ATMOS) PI on archiving the data and documentation.
- GES DISC received the original Stratospheric and Mesospheric Sounder (SAMS) and Selective Chopper Radiometer (SCR) data from the European Centre for Medium-Range Weather Forecasts (ECMWF), which are being compared with the SAMS and SCR data in the archive.
- Completed writing code and processing data for the first of the Nimbus 7 Limb Infrared Monitor of the Stratosphere (LIMS) datasets. The staff began writing code for the second LIMS dataset. Need to verify the results before it can be made available to the public.
- Staff continued to write documentation on the data recovery and code writing process.
- Data Available to the Public to date: 104,764 files from 6,680 tapes
 - Nimbus 1 HRIR (217 files from 10 tapes)
 - Nimbus 2 HRIR (2537 files from 1817 tapes)
 - Nimbus 2 MRIR (1616 files from 16 tapes)
 - Nimbus 3 HRIR (1278 files from 1012 tapes)
 - Nimbus 3 MRIR (2407 files from 20 tapes)
 - Nimbus 4 THIR 11.5 (1834 files from 1275 tapes)
 - Nimbus 4 THIR 6.7 (1540 files from 1016 tapes)
 - Nimbus 4 IRIS heritage data (214 files)
 - Nimbus 4 BUV CPOZ (2026 files from 4 tapes)
 - Nimbus 4 BUV PDB (12,434 files from 48 tapes)
 - Nimbus 4 BUV DCM (590 files from 3 tapes)
 - Nimbus 4 BUV DCW (368 files from 3 tapes)
 - Nimbus 4 BUV Radiance (12,174 files from 16 tapes)
 - Nimbus 4 Two BUV heritage datasets (84 files and 12,084 files)
 - Nimbus 5 THIR 11.5 (2,562 files from 104 tapes)
 - Nimbus 5 THIR 6.7 (86 files from 5 tapes)
 - Nimbus 5 ESMR (13,543 files from 166 tapes)
 - Nimbus 6 THIR 11.5 (469 files from 34 tapes)
 - Nimbus 6 THIR 6.7 (no tapes processed yet but 2 files were from an 11.5 tape)
 - Nimbus 6 SCAMS (4,052 files from 46 tapes)
 - Nimbus 6 HIRS (559 files from 86 tapes)
 - Nimbus 7 THIR CLDT (30,572 files from 595 tapes)
 - Nimbus 7 THIR BCLT (1516 files from 404 tapes)



Earth Science Data Recovery Work Remaining

Tapes recovered but not processed: 4,018 TAP Files

- To be processed by GES DISC: 73 TAP files
 - SME 1 TAP files
 - TIROS 63 TAP files
 - EOLE 1 2 TAP files
 - Explorer 7 4 TAP files
 - GEOS 2 TAP files
 - STS 51B 1 TAP files
- Sent to ASDC for processing: 634 TAP files
 - Nimbus 7 ERB 615 TAP files
 - Nimbus 7 SAM-II 19 TAP files
- TBD location for processing: 3,311 TAP files
 - Nimbus 7 SMMR 784 TAP files
 - Nimbus 7 SBUV/TOMS 883 TAP files
 - DMSP 84 TAP files
 - ATS-6 118 TAP files
 - SMS 1433 TAP files
 - GOES (labeled as Nimbus) 3 TAP files
 - STS 2 4 TAP files
 - STS 41G 2 TAP files
- Tapes Processed but Data Not Yet Validated: 38,153 files from 592 tapes
 - Nimbus 3 SIRS (x files from 6 tapes)
 - Nimbus 4 SIRS (y files from 3 tapes)
 - Nimbus 5 ITPR (z files from 2 tapes)
 - Nimbus 4 SCR (19,226 files from 28 tapes)
 - Nimbus 5 SCR (9481 files from 16 tapes)
 - Nimbus 6 LRIR (336 files from 6 tapes)
 - Nimbus 6 Merged Retrieval (1277 files from 86 tapes)
 - Nimbus 7 THIR NCLE (2180 files from 16 tapes)
 - Nimbus 7 SAMS Grid-T (1541 files from 8 tapes)
 - Nimbus 7 SAMS RAT (1253 files from 211 tapes)
 - Nimbus 7 SAMS ZMT (2 files from 2 tapes)
 - Nimbus 7 LIMS RAT (2857 files from 219 tapes)

Tapes not yet recovered (that we know of): 1,398 7-track; 2,452 9-track; 1,428 3480-tape

- To be recovered by GES DISC: 1,360 7-track
 - N2 HRIR: 511 7-track
 - N3 HRIR: 26 7-track
 - N4 THIR 6.7: 8 7-track
 - N5 SCR: 10 7-track
 - N5 SCMR 5 7-track
 - N5 ESMR 40 7-track
 - N5 THIR 6.7 636 7-track
 - N6 THIR 6.7 111 7-track
 - N6 THIR 11.5 2 7-track
 - N7 THIR BCLT 3 7-track
 - N7 THIR CLDT 1 7-track
 - N7 ERB 7 7-track
- To be recovered by TBD others: 38 7-track; 881 9-track; 77 3480-tape
 - N7 ERB: 731 9-track, 77 3480-tape
 - N7 SAM-II: 150 9-track
 - SMS: 38 7-track
- To be determined for the need for recovery: 1,571 9-track; 1351 3480-tape
 - N7 SMMR: 623 9-track
 - SME: 34 9-track
 - GOES 1,2,3: 647 9-track, 1351 3480-tape
 - STS 2: 16 9-track
 - STS 41G: 251 9-track